DOCUMENT RESUME

ED 223 249 IR 010 518

AUTHOR Brannstrom, Lauritz
TITLE Evaluating Visual Displays by Means of Judgment and Reproduction Methods. Umea Psychological Reports No. 152.
INSTITUTION Umea Univ. (Sweden).
SPONS AGENCY Swedish Council for Social Science Research, Stockholm.
PUB DATE 80
NOTE 13p.; For related documents, see IR 010 515-517.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cognitive Processes; College Students; *Evaluation Methods; *Patterned Responses; Psychological Studies; *Research Methodology; *Visual Measures; *Visual Perception; Visual Stimuli
IDENTIFIERS Form Reproduction Task; Magnitude Estimations; Paper and Pencil Tests; Pattern Goodness; Selective Attention; *Stimulus Characteristics

ABSTRACT

The spatial layout of visual displays was evaluated by means of three magnitude estimation tasks and one reproduction task. Each stimulus display consisted of 16 randomly-drawn, two-digit numbers. Six different spatial layouts, or display types, were used: systematic vs. random spatial order of numbers; regular vs. irregular boundaries of the stimulus pattern; and spatially grouped vs. ungrouped numbers, representing spatial order to different extents. A videotape recorder and television monitor were used for stimulus display presentations, with subjects seated at desks in a shielded room in front of the monitors. Presentation time for each stimulus display was 1.25 seconds. Decreased redundancy of the spatial configuration of elements was reflected in lower estimations and reproduction scores. The tasks referring to visual aspects were quite sensitive to variations in spatial configuration, but even the tasks referring to content attributes were affected in a similar way. The ease of application, the reliability, and the sensitivity of the paper and pencil methods make them suitable for use to complement operational tests of visual displays. Ten references are provided. (Author/LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223249

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EVALUATING VISUAL DISPLAYS BY MEANS OF JUDGMENT AND REPRODUCTION METHODS

by

Lauritz Brannstrom

UMEA PSYCHOLOGICAL REPORTS

No. 152 1980

Department of Psychology
University of Umea

IR 010518

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Lauritz Brannstrom

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EVALUATING VISUAL DISPLAYS BY MEANS OF JUDGMENT AND REPRODUCTION METHODS

Brännström, L. Evaluating visual displays by means of judgment and reproduction methods. Umeå Psychological Reports, No. 152, 1980. - The spatial layout of visual displays was evaluated by means of three magnitude estimation tasks and one reproduction task. Decreased redundancy of the spatial configuration of elements was reflected in lower estimations and reproductions scores. The tasks referring to visual aspects were quite sensitive to variations in spatial configuration but even the tasks referring to content attributes were affected in a similar way. The ease of application, the reliability and the sensitivity of the paper and pencil methods motivate their use as complementary methods to operational tests of visual displays.

The transfer of information from visual displays to the human user can be studied in different ways: by paper and pencil methods, operational tests, evaluations by expert judgment and through simulation (Crook, Raben & Wade, 1967). Some of the merits of paper and pencil methods are the ease with which they can be used and their low cost. But most important is that they can be used for preventing fallacies in the design of how information should be arranged in e.g. tables, lists and visual display terminals. The main disadvantage is the "off-line" character of the judgment methods as compared to real operational tests. The aim of the present study was to use rating and reproduction methods for evaluating the importance of spatial layout factors in visual displays.

Spatial layout seems to be closely correlated to pattern goodness, i.e. the structural redundancy of spatial patterns (Garner, 1974). The positive effects of pattern goodness on different information processing tasks are well-documented (e.g. Bell & Handel, 1976; Checkosky & Whitlock, 1973; Clement, 1964; Clement & Varnadoe, 1967; Garner & Sutliff, 1974). Glushko (1975) found by using multidimensional scaling and hierarchical clustering analysis that judgments of pattern goodness by the method of pair comparisons best could be described by a two-dimensional representation of the derived pattern space: Redundancy-

randomness being the primary dimension and compactness-dispersion being the second one. Visual displays should then be more easily processed when display elements form a good spatial pattern.

The spatial layouts used in the present study varied along three dimensions, namely systematic vs. random spatial order of elements, regular vs. irregular boundaries of the stimulus pattern and spatially grouped vs. ungrouped elements. The stimulus material was taken from a detection experiment by Brännström (1980) which evaluated the effects of spatial configuration on search performance. These configurations could be rank ordered with respect to the degree of spatial redundancy. The more redundant the spatial configuration the more positively it should be rated and reproduced.

The importance of spatial order depends on the nature of the operational task. This dependency should also be reflected in judgments of visual displays. Thus, judgments of visual form should be more sensitive to spatial order than judgments of the informational content. Two tasks were used which referred to visual aspects: (1) reproduction of the patterns, and (2) judgments of the ease of detecting and remembering spatial patterns. The other two tasks mainly referred to the meaning of the display elements: Subjects were to judge the ease of (3) detecging a target number among a set of simultaneously displayed two-digit numbers, and (4) of detecting numerical orderings of the numbers within the display.

In the experiment by Brännström (1980) the choice of a visual scanning strategy depended on the spatial configuration of elements. When the configuration was redundant (systematic) the displays were searched in an ordered fashion (from upper left to lower right). This search affected performance positively when also the numerical order was redundant but not when it was random. On the other hand, when the spatial configuration was random the scanning seemed to depend mainly on the visual clarity of the target element; the informational content did not affect performance in that condition. Thus, an interesting question is whether the paper and pencil methods used here will reflect this conditional choice of scanning strategy.

Stimulus material

Each stimulus display consisted of sixteen two-digit numbers randomly drawn without replacement from the set of two-digit numbers. (Numbers ending on zero or those containing identical pairs of digits were excluded). Six different spatial layouts (display types) were used. The display types varied on the following dimensions: (a) systematic vs. random spatial order of numbers, (b) regular vs. irregular boundaries of the stimulus pattern and (c) spatially grouped vs. ungrouped numbers. Thus, the generated display types represented spatial order to different extents, from systematic to random spatial configurations. Display type I was a redundant (systematic) pattern of equally spaced numbers ordered in rows and columns (Figure 1 a). Display type II-VI were transformations of type I. Type II and III were spatially grouped; type III differing from II only with respect to one inner position of elements (Figure 1 b and c). Type IV had inner positions that were determined at random (Figure 1 d) while type V and VI were random display types (Figure 1 e and f); V was spatially grouped while VI was ungrouped.

Figure 1 a-f. Display types.

The rank order of displays with respect to spatial redundancy should thus be from I to VI.

For the judgment tasks two stimulus displays of type I and II were generated by means of an alfa-numeric computer display. For each of the types III-VI eight stimulus displays were generated with each stimulus display replicated twice. Thus, there were totally 68 stimulus displays. These were recorded in random order on videotape. In the reproduction task the same display types were used as in the judgment tasks but with only two replicates of each; this resulted in 12 stimulus displays. For practice trials, six different stimulus displays were generated.

Procedure

A videotape recorder and a 26" TV-monitor were used for presentation of the stimulus displays. Subjects were sitting at desks in a shielded room in front of the TV-monitor. At the most five subjects participated simultaneously in an experimental session. The distances of the subjects to the TV-monitor varied between 2.0 to 3.5 meters, and the stimulus display had a width of .25 m and a height of .15 m. Thus, the display width could range from 7.2^{o} to 4.1^{o} of visual angle while the display height could range from 4.3^{o} to 2.5^{o} of visual angle. The luminance level corresponded to that used in normal TV-viewing, i.e. the room was sparsely illuminated. The two-digit numbers in the stimulus displays appeared as bright figures against a dark-grey background.

The presentation time for each stimulus display was 1.25 sec. In the judgment tasks there was a pause of 5.75 sec for filling in the response sheet. The total time for the three judgment tasks was 40 minutes. In the reproduction task the subjects were allowed 2 minutes for reproducing a display on a response matrix. The empty response matrix mapped the positions of the display. There were 140 blank positions and the subject should fill 32 of them with x:es instead of numbers; each pair of x:es corresponding to a two-digit number. The reproduction task was completed within about 45 minutes.

In the judgment tasks the subjects were instructed to make magnitude estimations. The judgment attributes were: (a) ease of detecting a target number within the display. The magnitude scale extended from 0 (impossible to detect the number) to 10 (should almost always find the number), (b) ease of detecting and remembering spatial patterns (0 - impossible; 10 - very easy to detect and remember the spatial

pattern), (c) ease of detecting numerical order (0 - impossible; 10 - should almost always detect numerical order).

Seven female and six male university students served as voluntary subjects in one experimental session lasting approximately two hours.

Results

Overall reliability coefficients were computed for each task by inter-correlating the magnitude estimations for the first and second replicate across stimulus displays. In the reproduction task a correct score was obtained if a marked x also corresponded to an actual digit position on the stimulus display. Marking either one or two of the digit positions correctly for a displayed number yielded the same scoring. The corrected split-half reliability coefficients are summarized in Table I.

Table I. Split-half reliabilites for the magnitude estimation (N=13) and reproduction (N=12) tasks.

TASK	RELIABILITY
ESTIMATION	
(a) ease of detecting a target number	.845
(b) ease of detecting and remembering spatial patterns	.893
(c) ease of detecting numerical order	.904
REPRODUCTION	.923

The reliability coefficients are high. Task order probably had a slight influence on the experiment, with the result that the later judgment tasks yielded more stable estimations due to increased experience to the stimulus material.

Figure 2 shows the mean estimations and the mean correct reproductions as a function of the assumed rank order of displays with respect to spatial order.

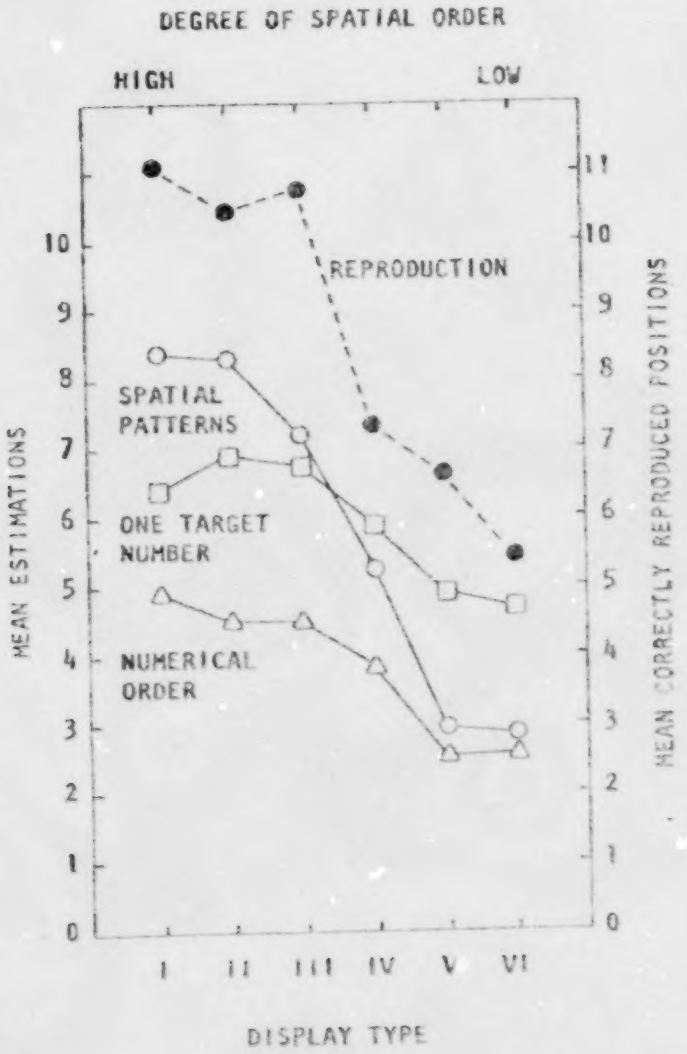

Figure 2. Mean ratings and mean correct reproduction scores as a function of spatial order (i.e. display type).

It is obvious that the spatial characteristics of the displays affect mean estimations and mean reproduction scores in the predicted way. An ANOVA of the estimation data yielded significant main effects of tasks and display types ($F(2,24) = 7.593$; $p < .01$ and $F(5,60) = 9.761$; $p < .01$).

The estimation of spatial patterns (e.g. visual aspects) is more sensitive to spatial order than the estimations referring to the informational content of the displays. The interaction between judgment task and display type was significant ($F(10,120) = 6.483$; $p < .01$). However, the still marked effect of spatial order on the mean estimations referring to the informational content suggests a dependency between spatial layout and content. For the reproduction task there is a statistically significant effect of display type ($F(5,55) = 12.405$; $p < .01$).

The effects of spatial order on mean estimations, mean correct reproductions and standard deviations in the dependent variables are summarized in Table 2.

Table 2. Mean estimations, mean correct reproductions and standard deviations for the different display types. Predicted and obtained rank orders with respect to degree of spatial redundancy of element positions are given within parentheses.

DISPLAY TYPE (Predicted rank within parenthesis)		ESTIMATION: Ease of detecting One number (Obtained ranks within parentheses)	Spatial patterns	Numerical order	REPRODUCTION
I (1)	$\bar{x}$	6.34 (3)	8.39 (1)	4.89 (1)	11.09 (1)
	s	3.07	3.50	4.21	4.23
II (2)	$\bar{x}$	6.92 (1)	8.31 (2)	4.54 (2)	10.46 (3)
	s	2.74	3.15	3.84	4.21
III (3)	$\bar{x}$	6.76 (2)	7.18 (3)	4.49 (3)	10.79 (2)
	s	2.80	2.82	3.66	4.20
IV (4)	$\bar{x}$	5.91 (4)	5.28 (4)	3.87 (4)	7.33 (4)
	s	2.27	2.35	3.12	2.75
V (5)	$\bar{x}$	4.93 (5)	2.94 (5)	2.52 (6)	6.63 (5)
	s	2.57	2.35	2.49	2.48
VI (6)	$\bar{x}$	4.66 (6)	2.84 (6)	2.56 (5)	5.46 (6)
	s	2.59	2.54	2.30	1.91

The rank order of display types based on the data in Table 2 accords well with the predicted rank order. Deviations from predicted ranks are found for display types I, II and III but these deviations are probably due to the fact that they are quite similar as regards the degree of spatial redundancy of element positions. Standard deviations of estimations and correct reproductions tend to decrease with lower ranks of spatial redundancy (predicted ranks). The effect depends either on a greater consistency of responses, or a restriction of range in performance.

In Table 3 the overall correlations between dependent measures in the different tasks are presented.

Table 3. Overall correlations between tasks across display types; partial correlations are given within parentheses.

	ESTIMATION		
	Ease of detecting		REPRODUCTION
	Spatial patterns	Numerical order	
Ease of detecting			
One number	.65 (.54)	.45 (.27)	-.08
Spatial patterns		.52 (.49)	.19
Number order			.05

The judgment tasks are positively intercorrelated. The reproduction data do not correlate significantly with any of the judgment data, a result probably attributable to the subjects' ability to make reproductions. Thus, the sensitivity of the reproduction measure to spatial order depends on a few sophisticated subjects whose performance affects the mean scores heavily.

Discussion

The consistency of the paper and pencil methods used in the present study is confirmed by the high reliability coefficients obtained. This fact means that the methods are suitable for practical application as regards reliability.

The obtained rank order of display types based on mean estimations and mean correct reproductions corresponded well to the predicted rank order of the display types with respect to the spatial redundancy of element positions. Thus, the methods are sensitive to variations in the spatial order of elements within visual displays; the more redundant the spatial pattern is, the higher are the mean estimations and the mean reproduction scores.

The evaluation process seems to be very sophisticated. This is reflected in the clear interaction effect between display type and judgment task. Estimations referring to visual attributes of the display are more sensitive to variations of spatial order than estimations referring to attributes concerning the informational content of the elements. However, the still marked effect of spatial order on the two judgment tasks referring to content attributes of the display elements suggest a dependency between spatial layout and content. This result is in accordance with the type of dependency found in the experiment by Brännström (1980) which involved operational tests of the effect of spatial factors on the choice of visual scanning strategy.

In that experiment an ordered scanning was observed for display types with a redundant spatial order of elements. For the random display used (i.e. type VI) subjects scanned according to clarity of elements, i.e. targets near the line of sight or targets at the boundaries of the display pattern were detected more frequently than those at other locations. The ordered search was efficient only when both spatial and numerical order were systematical but inefficient when the numerical

order of elements was random. For the random display scanning was unaffected by numerical order. In the present experiment no such conditional dependency between the spatial configuration and the content of the elements is revealed by the estimations concerning ease of detecting a target number. One reason could be that subjects were not made explicitly aware of numerical order until the last judgment task. Thus, estimations of the ease of detecting a target number were made as if spatial and numerical order were independent. Another reason could be that the magnitude estimations are too crude to reflect such intricate dependencies. However, the impact of spatial order on the estimations of the informational content of elements is unquestionable.

The data from the reproduction task show that the method is quite sensitive to variations in spatial order of elements. The conspicuous similarity of the curves for mean correct reproductions and mean estimations of the ease of detecting and remembering spatial patterns lends further support to the importance of configurational aspects. The only disadvantage of the reproduction task is that some subjects will not meet the rather difficult task requirements; a critique not applicable for any of the judgment tasks.

From the present study it can be concluded that visual displays can be reliably evaluated by means of judgment and reproduction methods. The methods are capable of revealing many, but not all, of the facets found with operational tasks. Thus, display evaluation by means of paper and pencil methods should be used with some care and taken as "symptomatic indicators of areas for needed improvement" (Siegel et al., 1975; p. 287).

This investigation was supported by a grant from the Swedish Council for Social Science Research.

References

Bell, H.H., & Handel, S. The role of pattern goodness in the reproduction of backward masked patterns. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 139-150.

Brännström, L. Effects of spatial configuration on search of visual displays. Umeå Psychological Reports, No. 151, 1980. University of Umeå.

Checkosky, S.F., & Whitlock, D. Effects of pattern goodness on recognition time in a memory search task. Journal of Experimental Psychology, 1973, 100, 341-348.

Clement, D.E. Uncertainty and latency of verbal naming responses as correlates of pattern goodness. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 150-157.

Clement, D.E., & Varnadoe, K.W. Pattern uncertainty and the discrimination of visual patterns. Perception & Psychophysics, 1967, 2, 427-431.

Crook, M.N., Raben, M.W., & Wade, E.A. Trends and developments in visual displays. Medford, Mass.: Tufts University Institute for Psychological Research, Human Engineering Information and Analysis Service, 1967.

Garner, W.R. The processing of information and structure. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Garner, W.R., & Sutliff, D. The effect of goodness on encoding time in visual pattern discrimination. Perception & Psychophysics, 1974, 16, 425-430.

Glushko, R.J. Pattern goodness and redundancy: Multidimensional scaling and hierarchical clustering analysis. Perception & Psychophysics. 1975, 17, 158-162.

Siegel, A.I., Fischl, M.A., Macpherson, D. The analytic profile system (APS) for evaluating visual displays. Human Factors, 1975, 17, 278-288.